

## Interface de Recueil de Données Imparfaites pour le CrowdSourcing

Constance Thierry, Géry Casiez, Jean-Christophe Dubois, Yolande Le Gall,  
Sylvain Malacria, Arnaud Martin, Thomas Pietrzak, Pierrick Uro

### ► To cite this version:

Constance Thierry, Géry Casiez, Jean-Christophe Dubois, Yolande Le Gall, Sylvain Malacria, et al..  
Interface de Recueil de Données Imparfaites pour le CrowdSourcing. EGC 2020 - Humains et IA,  
travailler en intelligence Atelier de la conférence, Jan 2020, Bruxelles, Belgique. hal-02465761

**HAL Id: hal-02465761**

**<https://hal.inria.fr/hal-02465761>**

Submitted on 4 Feb 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Interface de Recueil de Données Imparfaites pour le CrowdSourcing

Constance Thierry\*, G ry Casiez\*\*, Jean-Christophe Dubois\*, Yolande Le Gall\*, Sylvain  
Malacria\*\*, Arnaud Martin\*, Thomas Pietrzak\*\*, Pierrick Uro\*\*

\*Univ Rennes, CNRS, IRISA, DRUID

prenom.nom@irisa.fr,  
<http://www-druid.irisa.fr>

\*\*Inria Lille - Nord Europe, CRISTAL - UMR 918

prenom.nom@inria.fr,  
<https://loki.lille.inria.fr/>

**R sum .** Le *crowdsourcing* se d finit par l'externalisation de t ches   une foule de contributeurs. La foule, tr s diversifi e sur ces plateformes, inclut des contributeurs de niveaux de qualification variables et qui r alisent la t che de mani re plus ou moins consciencieuse. Cette diversit  de profils induit des contributions de qualit  in gale. Afin de palier ce probl me nous  tudions une interface utilisateur offrant aux contributeurs une plus grande capacit  d'expression dans leurs r ponses. Elle permet aux contributeurs de fournir une r ponse impr cise en cas d'h sitation ainsi que de renseigner leur confiance dans leurs r ponses. Cette interface a  t  d ploy e en demandant aux utilisateurs de r pondre   des questions de difficult  variable et dont les r ponses sont connues. Deux plages de difficult  ont  t  test es et nos r sultats sugg rent que la certitude report e par les utilisateurs est bien fonction de la difficult  absolue de la question, et non pas r partie de mani re relative sur la plage de difficult .

## 1 Introduction

Aujourd'hui encore, certaines activit s ne peuvent  tre r alis es par ordinateur car elles sont trop co teuses en temps ou n cessitent une expertise humaine. Le *crowdsourcing* permet d'externaliser ces t ches sur des plateformes d di es afin qu'elles soient accomplies par une multitude d'utilisateurs, appel s contributeurs. Les donn es recueillies issues de sources h t rog nes sont ensuite fusionn es pour en extraire une information fiable et permettre une prise de d cision. Cette pratique offre aux employeurs l'avantage de traiter une t che dans de courts d lais et avec un co t r duit. Cependant une probl matique du domaine est d'obtenir des donn es fiables et pertinentes. En effet, les contributions humaines de qualit  in gale, comportent des imperfections inh rentes   toute  valuation subjective, parfois amplifi es par un manque d'expertise ou de s rieux, mais aussi li es au mode de recueil des donn es. Ainsi, alors que les contributeurs sont fr quemment confront s   des situations d'ind cision, les possibilit s d'expression des interfaces de recueil d'informations ne permettent pas de formuler finement

les réponses ni d'indiquer la validité accordée. Nos travaux ont pour objectif de proposer une interface dédiée au *crowdsourcing* permettant aux contributeurs d'exprimer des réponses plus riches comportant des imperfections exploitables dans le processus de décision à l'aide de théories de l'incertain. A l'instar de Koulougli et al. (2016), Ouni et al. (2017), Abassi et Boukhris (2018), Ben Rjab et al. (2016) nous avons opté pour une modélisation fondée sur les fonctions de croyance.

La section 2 présente l'interface proposée et le protocole expérimental suivi, la section 3 présente une analyse des résultats obtenus. Puis, la section 4 compare une modélisation des réponses par la théorie des fonctions de croyance avec une modélisation probabiliste et un vote majoritaire. Enfin nous concluons cette étude et présentons nos perspectives de travail.

## 2 Interface et protocole expérimental

L'interface étudiée a pour objectif d'offrir au contributeur la possibilité d'exprimer une réponse imprécise, en lui permettant de donner plusieurs choix en cas d'hésitation, mais aussi incertaine en attribuant un niveau de confiance en sa réponse. Afin de valider cette interface, 3 phases expérimentales sont considérées afin d'analyser distinctement le comportement du contributeur confronté à des situations d'incertitude puis d'indécision et enfin aux deux simultanément. Nous aborderons dans cet article uniquement la première phase expérimentale visant à valider qu'il y a bien une corrélation entre la facilité de réalisation d'une tâche et la certitude exprimée par le contributeur en s'appuyant sur une courbe psychométrique. Cette campagne de *crowdsourcing* a été réalisée sur la plateforme Figure Eight<sup>1</sup>.

La tâche repose sur la perception visuelle, ce qui nous permet d'avoir une vérité terrain, de contrôler de manière continue la difficulté de la tâche liée à la différence de longueur entre les segments et de ne pas nécessiter de qualification spécifique de la part du contributeur. Dans cette expérimentation dont l'interface est présentée figure 1, il est demandé au contributeur de désigner le plus grand segment parmi un ensemble de 5 segments, en renseignant la certitude dans sa réponse. La réponse correcte est incluse dans les 5 propositions, ce qui donne un taux de chance de 20 %. La *position* de la bonne réponse pouvant influencer les résultats, chaque question est dupliquée pour les 5 positions possibles.

L'échelle de certitude présentée au contributeur est la suivante : totalement incertain (0), incertain (1), peu incertain (2), ni certain ni incertain (3), peu certain (4), certain (5), totalement certain (6). Il est notifié au contributeur au début de la campagne qu'il a le droit d'être incertain dans sa réponse et que ceci n'aura pas d'impact lors de sa rémunération.

La difficulté de cette tâche suit une loi de Weber, c'est-à-dire que le rapport entre le seuil de perception  $\Delta_I$  et la valeur témoin  $I$  est une constante. La longueur témoin a été déterminée empiriquement à 40 mm. La difficulté variait selon la valeur de  $\Delta$  entre la bonne réponse et les bâtons témoins. Nous avons mené des études empiriques en testant notre interface en laboratoire afin de déterminer  $\Delta_I$ .

Lors de ces études préliminaires, nous avons observé que les participants estimaient leur certitude de manière relative, c'est-à-dire qu'ils donnaient une forte certitude pour la valeur de  $\Delta$  la plus élevée, et que les valeurs de certitude intermédiaires étaient réparties entre les valeurs de  $\Delta$  extrêmes (Figure 3, courbes Labo). Néanmoins les participants ne parviennent

---

1. <https://www.figure-eight.com>

FIG. 1 – Interface proposée aux contributeurs.

pas à assigner les valeurs extrêmes de certitude à leurs réponses, par exemple si ces certitudes ont des valeurs de 0 et 5, alors les participants indiqueront généralement leur certitude entre 1 et 4. Nous avons donc choisi d'expérimenter deux séries  $\Delta_0$  et  $\Delta_1$  afin de vérifier cette hypothèse. Ces deux séries ont été effectuées par des participants différents afin d'éviter un effet d'apprentissage qui pourrait biaiser l'expérience. La valeur de  $\Delta$  la plus basse est de 0. Le taux d'erreur pour cette valeur devra être proche du taux de chance. La valeur de  $\Delta$  maximale est choisie afin d'être clairement identifiable selon nos études préliminaires. Les deux séries comportaient 6 valeurs, un compromis entre précision de mesure et durée de l'expérience. Les séries contiennent plus de valeurs autour du seuil de perception obtenu lors des études préliminaires ( $300 \mu m \leq \Delta_I \leq 600 \mu m$ ) :

$$\Delta_0 = \{0, 300, 600, 900, 1200, 1500\}$$

$$\Delta_1 = \{0, 300, 600, 1200, 1800, 2400\}$$

Le protocole expérimental est donc  $5 \text{ positions} \times 6 \text{ valeurs} \times 3 \text{ répétitions} = 90$  essais par participant. La série de valeurs est un facteur intra-sujet, balancé entre les utilisateurs. Il était de plus demandé aux participants d'installer le logiciel *pointing server*<sup>2</sup> afin de détecter la taille de l'écran du contributeur. Au total, 118 contributeurs ont réalisé la campagne de *crowdsourcing*, 63 pour  $\Delta_0$  et 55 pour  $\Delta_1$ , la base de données des réponses considérées contient ainsi 5670 contributions pour  $\Delta_0$  et 4950 pour  $\Delta_1$ . L'étude préliminaire réalisée en laboratoire incluant 6 contributeurs pour  $\Delta_0$  et 5 pour  $\Delta_1$  est également étudiée par la suite.

En fin d'expérimentation, plusieurs questions permettant d'avoir un retour sur l'expérience ont été introduites. Elles sont détaillées dans la section présentant les résultats.

2. <https://github.com/INRIA/libpointing>

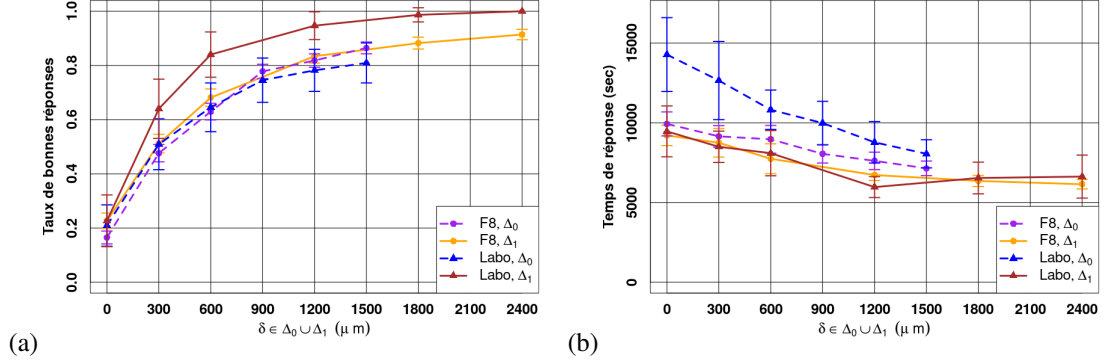


FIG. 2 – Taux de bonnes réponses (a) et temps de réponse (b) des contributeurs.

### 3 Analyse des résultats

Durant la campagne de crowdsourcing 118 contributeurs ont réalisé intégralement les tests. 20 contributeurs supplémentaires ont débuté l'expérimentation mais ne l'ont pas menée à terme. La figure 2(a) représente le taux de bonnes réponses de l'ensemble des contributeurs pour chaque degré de difficulté  $\delta$ . Plus  $\delta$  est grand, plus la différence de taille entre le plus grand segment et les autres est importante et donc plus la tâche est facile pour le contributeur. Conformément à ce qui est attendu, le taux de bonnes réponses des contributeurs augmente bien avec  $\delta$ . De plus, les taux de bonnes réponses pour  $\Delta_0$  et  $\Delta_1$  sont très proches pour les données de la campagne de crowdsourcing et celles de laboratoire, ce qui montre que la difficulté de la tâche est bien perçue par le contributeur quelque soit l'ensemble de difficulté  $\Delta$  choisi. — Nous déterminons à cet effet la courbe psychométrique des utilisateurs présentant la distribution des certitudes renseignées selon une difficulté croissante de la tâche. —

Considérons maintenant le temps de réponse du contributeur à la tâche. La figure 2(b) montre une diminution du temps de réponse moyen des contributeurs avec l'augmentation de  $\delta$ . Ainsi, plus la tâche est facile, moins les contributeurs ont besoin de temps pour répondre car ils perçoivent rapidement quel est le segment le plus grand. Une autre constatation pour cette figure, est l'écart de temps de réponse entre les contributeurs. Les contributeurs ayant réalisé l'expérience avec  $\Delta_0$  ont eu besoin de davantage de temps pour répondre en comparaison aux contributeurs ayant réalisé l'expérience pour  $\Delta_1$ . Cette différence de temps peut s'expliquer par le fait que l'écart des valeurs de  $\delta \in \Delta_1$  est plus important. Par conséquent, le segment le plus long est plus rapidement perceptible pour les grandes valeurs de  $\delta$  et induit davantage de rapidité pour les plus faibles valeurs. La difficulté générale de la campagne a donc un fort impact sur le temps de réponse des contributeurs sur ces plateformes.

Nous observons (figure 3) qu'en moyenne, les contributeurs de la plateforme de crowdsourcing FigureEight sont assez certains de leurs réponses, et ce même lorsque la tâche est la plus difficile et que tous les segments sont de longueurs identiques ( $\delta = 0$ ). De manière positive, la diminution de la difficulté de la tâche avec l'augmentation de  $\delta$  s'accompagne d'une

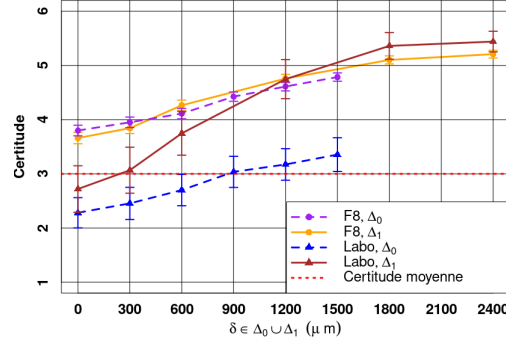


FIG. 3 – **Certitude moyenne** des contributeurs en fonction de la difficulté de la tâche.

augmentation de la certitude des contributeurs. Néanmoins, une diminution plus importante de la certitude lorsque la difficulté de la tâche augmente était attendue. En effet, lorsque tous les segments sont de taille égale ( $\delta = 0$ ), il n'est pas possible pour le contributeur de sélectionner le bon segment. Dans ces conditions, le contributeur est alors contraint de choisir un segment au hasard, il devrait alors spécifier qu'il est totalement incertain (0). L'évolution de la certitude d'après la difficulté des questions pour les données de laboratoire avec  $\Delta_1$  est davantage en accord avec nos attentes. Cette difficulté de la part du contributeur de la plateforme de crowdsourcing de spécifier qu'il est incertain peut être due au fait qu'il n'ose pas affirmer cette incertitude de peur que son travail soit déprécié même si dans cette expérience, une contribution n'est pas dévalorisée en raison d'une forte incertitude. Au contraire, il est attendu de la part du contributeur que sa certitude en sa réponse varie suivant la difficulté de la question. Ainsi, un contributeur dont la certitude est constante est moins crédible qu'un contributeur qui reconnaît que sa certitude change suivant la difficulté de la question.

Ce dernier point est abordé dans la section suivante.

Un autre fait intéressant est l'évolution du taux de bonnes réponses des contributeurs au cours de la tâche. Afin de modéliser cette évolution le taux de bonnes réponses est mesuré pour une question  $q$  de la manière suivante :

$$TBR(q) = \sum_{c \in C} \sum_{i \in [1, q]} \frac{\text{isGoodAnswer}_{c,i}}{q * |C|}$$

Dans l'équation ci-dessus,  $C$  est l'ensemble des contributeurs ayant réalisé la campagne, et le booléen  $\text{isGoodAnswer}_{c,i}$  est égal à 1 si le contributeur  $c$  a renseigné la bonne réponse à la question  $i$ . La figure 4 est la représentation graphique de l'évolution du taux de bonnes réponses des contributeurs au cours de la campagne pour  $\Delta_0$  et  $\Delta_1$ . Pour les contributeurs ayant réalisé la campagne avec  $\Delta_0$ , ce taux diminue faiblement au cours de l'expérience. Il semblerait que les contributeurs soient progressivement moins attentifs. Cette diminution de l'attention du contributeur peut s'expliquer par le fait que la campagne utilisant  $\Delta_0$  est plus complexe que celle avec  $\Delta_1$  et que la tâche devient fastidieuse au fur et à mesure de l'expé-

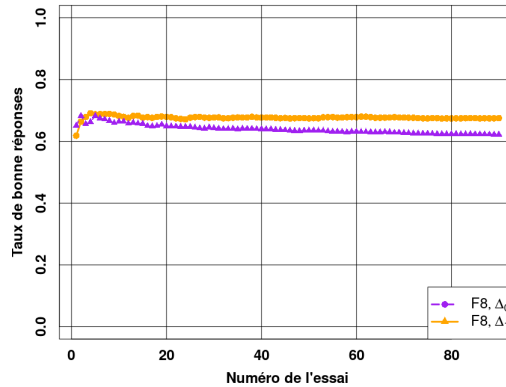


FIG. 4 – Taux de bonnes réponses des contributeurs au cours de la tâche pour  $\Delta_0$  et  $\Delta_1$ .

rience. Pour les contributeurs ayant réalisé la campagne de crowdsourcing pour  $\Delta_1$ , le taux de bonnes réponses varie un peu sur les premières questions puis est plus constant. Cette variation du taux sur les premières réponses est associé au temps d'apprentissage du contributeur, ce temps d'apprentissage est aussi visible pour la courbe associée à  $\Delta_0$  pour les premières questions.

Une fois la campagne réalisée par les contributeurs sur *Figure Eight*, il leur était demandé de répondre à un dernier questionnaire afin d'avoir leurs retours sur cette expérience. Les questions posées et réponses proposées sont les suivantes :

- (q1) Pour vous, renseigner le segment le plus long était :  
Très facile, facile, ni facile ni difficile, difficile, très difficile.
- (q2) Pour vous, exprimer votre certitude était :  
Très facile, facile, ni facile ni difficile, difficile, très difficile.
- (q3) Avez-vous hésité avant de répondre ?  
Presque toujours, souvent, occasionnellement, rarement, jamais.
- (q4) Dans le cas où vous hésitez souvent, quel en est la cause ?  
La tâche devient fastidieuse, vous avez un doute entre plusieurs segments, autre (à préciser).

Les pourcentages de contributeurs ayant répondu aux questions (q1) et (q2) sont illustrés figure 5, nous revenons sur ces résultats plus tard. A la question (q3), 15% des contributeurs ont répondu qu'ils hésitaient presque toujours, 25% souvent, 31% occasionnellement, 17% rarement et 13% considèrent ne jamais avoir hésité. Or, parmi l'ensemble des questions, à 5 reprises les segments étaient tous de même taille et il était donc impossible de ne jamais hésiter sur la réponse à renseigner. Pour la question (q4) seules les réponses des contributeurs qui admettent avoir douté sont considérées, soit 87% de la foule. La majorité de ces contributeurs (84%) reconnaissent que leur hésitation est liée au fait que la tâche devient fastidieuse, alors

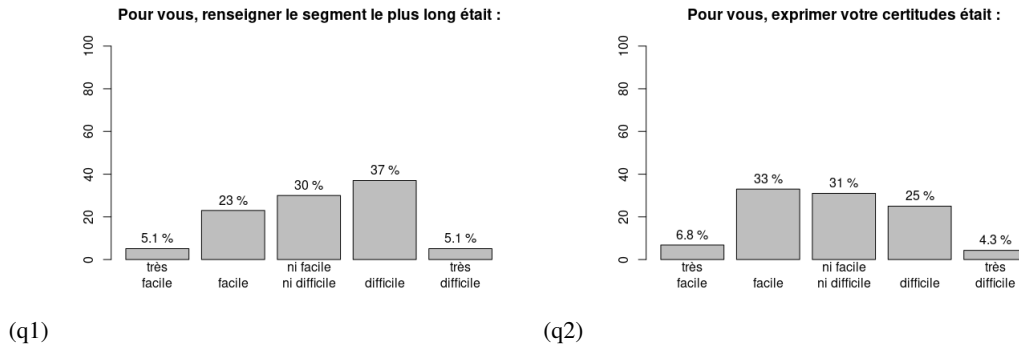


FIG. 5 – Pourcentage de contributeurs d'après leurs réponses aux questions (q1) et (q2).

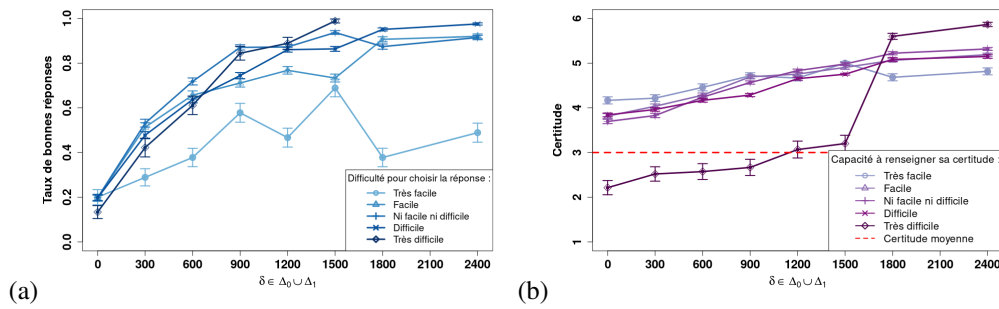


FIG. 6 – Taux de bonnes réponses et certitude pour 5 groupes de contributeurs.

que les 16% restants indiquent que cette hésitation est due à la taille des segments.

La figure 6(a) est une représentation graphique du taux de bonnes réponses en fonction de la difficulté  $\delta$  pour cinq groupes de contributeurs. Ces groupes sont définis d'après la réponse du contributeur à la question (q1). Le pourcentage de contributeurs appartenant à chaque groupe est donné par la figure 5(q1). Une part importante des contributeurs (37%) précise qu'il leur était difficile de renseigner parmi les 5 segments lequel était le plus long. Le groupe de contributeurs ayant spécifié qu'il était très facile de déterminer le segment le plus long est celui possédant le taux de bonnes réponses le plus faible. Plusieurs hypothèses peuvent être avancées telles qu'une mauvaise compréhension de l'expérience, ou encore un manque de sérieux qui les ont conduits à répondre aléatoirement à la campagne. Notons également que le groupe de contributeurs (5.1%) ayant reconnu que la tâche était très difficile est uniquement constitué de participants ayant réalisé la campagne pour  $\Delta_0$  ce qui confirme que  $\Delta_0$  est perçu comme plus difficile que  $\Delta_1$ .

Considérons maintenant l'évolution de la certitude moyenne de groupes de contributeurs constitués d'après leur réponse à la question (q2). Le pourcentage de contributeurs apparte-



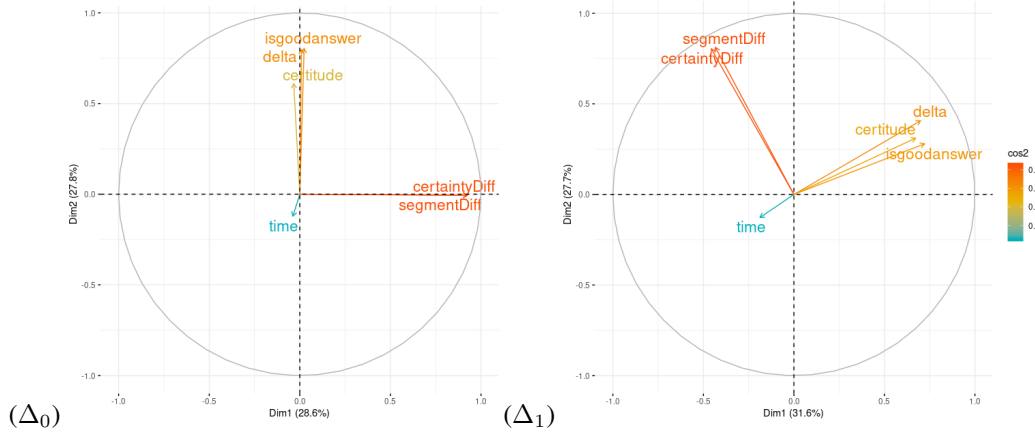


FIG. 7 – ACP : premier plan des variables.

nant à chaque groupe est donné par la figure 5(q2). Le groupe comportant le pourcentage le plus élevé de contributeurs (33%) est composé d'internautes ayant indiqué que renseigner leur certitude en leur réponse était facile. Néanmoins en observant sur la figure 6(b) la courbe associée à ce groupe, nous constatons que leur certitude varie peu avec la difficulté de la tâche. Les contributeurs considérant qu'il leur était très facile de spécifier leur difficulté sont pourtant ceux pour lesquels l'évolution de la certitude est la plus faible. Leur perception de la difficulté de la tâche n'est donc pas corrélée avec le niveau de confiance renseigné pour leurs réponses. Par ailleurs, ce sont les contributeurs pour lesquels il était très difficile de donner leur certitude qui ont fait évoluer ce niveau de certitude de façon la plus importante en fonction de la difficulté. En effet pour les questions les plus difficiles leur certitude moyenne est assez faible, et cette dernière devient réellement importante pour les questions les plus faciles ce qui était attendu des contributeurs.

Pour clore cette section, une analyse en composantes principales (ACP) a été réalisée pour  $\Delta_0$  et  $\Delta_1$  en considérant les paramètres suivants :

- delta : la difficulté  $\delta$  de la question
- isGoodAnswer : le booléen déterminant si la réponse du contributeur est juste
- certitude : certitude du contributeur en sa réponse
- time : le temps de réponse
- segmentDiff : la réponse à la question (q1)
- certaintyDiff : la réponse à la question (q2)

Considérons les graphiques des variables de ces ACP présentés figure 7. Le fait que le contributeur renseigne la bonne réponse (isGoodAnswer) à une question est fortement corrélé à la difficulté (valeur de delta) et à sa certitude que la réponse est juste. Ces variables sont en revanche totalement dé-corrélées de la perception que le contributeur a de sa capacité à renseigner le segment le plus long (segmentDiff) et de sa certitude (certaintyDiff) qui sont

corrélées entre elles. De plus la difficulté de la question est négativement corrélée au temps de réponse (time) du contributeur.

## 4 Modélisation des réponses

Dans cette section nous comparons 3 méthodes distinctes de modélisation et d'agrégation des réponses. La première utilise le vote majoritaire, la seconde une approche probabiliste et la troisième la théorie des fonctions de croyance.

**Vote majoritaire** Pour un vote majoritaire, les réponses des contributeurs sont modélisées par des fonctions indicatrices, telles que la réponse d'un contributeur  $c$  à la question  $q$  s'écrive  $r_{cq}$ . Soit  $\Omega$  l'ensemble des réponses possibles à la question  $q$ ,  $r_{cq}(X) = 1$ ,  $X \in \Omega$  si le contributeur choisit la réponse  $X$ , sinon  $r_{cq}(X) = 0$ . Les réponses sont agrégées en réalisant la somme des indicatrices, et la réponse ayant été donnée par le plus grand nombre de contributeurs est sélectionnée.

**Probabilité** On attribue à la réponse du contributeur la probabilité la plus haute, les autres éléments sont équiprobables avec  $P_{cq}(Y) \leq P_{cq}(X)$ .

$$\begin{cases} P_{cq}(X) = \alpha, X \in \Omega \\ P_{cq}(Y) = \frac{1-\alpha}{|\Omega|-1}, \forall Y \in \Omega \setminus X \end{cases} \quad (1)$$

La valeur de la probabilité minimale  $\alpha$  doit ici être supérieure ou égale à 0.2 pour que l'inégalité  $P_{cq}(Y) \leq P_{cq}(X)$  soit vérifiée, car  $|\Omega| = 5$ . En revanche avec  $\alpha = 0.2$  comme valeur numérique de l'incertitude totale du contributeur, nous obtenons une équiprobabilité totale représentant l'ignorance sur la réponse. Les valeurs numériques ( $\alpha$ ), associées à la certitude de la réponse du contributeur sont présentées dans le tableau 1.

**Fonctions de croyance** La théorie des fonctions de croyance a été introduite par Dempster (1967) et permet de modéliser l'incertitude et l'imprécision de sources d'informations imparfaites. Dans le cadre du *crowdsourcing*, les contributeurs sont des sources d'informations et leurs réponses dans nos expériences peuvent être imprécises et/ou incertaines. La théorie des fonctions de croyance considère un ensemble  $\Omega$  d'hypothèses, celui-ci est appelé cadre de discernement. Des fonctions de masses  $m : 2^\Omega \rightarrow [0, 1]$  sont alors définies telle que :

$$\sum_{X \in 2^\Omega} m(X) = 1 \quad (2)$$

L'ensemble des réponses proposées à un contributeur  $c$  pour une question  $q$  constitue le cadre de discernement  $\Omega$ . Il est alors possible d'associer à la réponse  $X$  du contributeur une fonction de masse à support simple :

$$\begin{cases} m_{cq}^\Omega(X) = \alpha \text{ with } X \in 2^\Omega \setminus \Omega \\ m_{cq}^\Omega(\Omega) = 1 - \alpha \end{cases} \quad (3)$$

Totalement incertain	Incertain	Un peu incertain	Ni certain, ni incertain	Un peu certain	Certain	Totalement certain
0.2	0.3	0.4	0.5	0.6	0.7	0.8

TAB. 1 – Valeur numériques  $\alpha$  associées à l'échelle de certitude

$\delta$	0	300	600	900	1200	1500
MV	0	15	15	15	15	15
Probabilité	0	15	15	15	15	15
Fonctions de croyances	0	15	15	15	15	15

TAB. 2 – Nombre de bonnes réponses (/15) pour  $\Delta_0$ 

Cette fonction caractérise le fait que le contributeur croit partiellement en la réponse  $X$  avec une masse  $\alpha$  mais pas davantage. L'ignorance est modélisée dans la théorie des fonctions de croyance par l'élément  $\Omega$ . Notons que la réponse  $X$  du contributeur peut être imprécise. Les valeurs numériques de  $\alpha$  utilisées sont les mêmes que pour notre modèle probabiliste (tableau 1).

**Tests réalisés** Une comparaison des trois méthodes énoncées ci-dessus est proposée pour l'agrégation des données de crowdsourcing. Pour ce faire, les réponses des contributeurs sont agrégées pour les 5 segments des 3 blocs pour chaque niveau de difficulté, soit un total de 15 valeurs agrégées par  $\delta$ . Les tableaux 2 et 3 renseignent le nombre de fois où les réponses agrégées sont correctes sur les 15 questions considérées. Pour ces deux tableaux, il est constaté que les trois méthodes proposées offrent les mêmes résultats pour l'agrégation des réponses. Ceci montre que la majorité des contributeurs étaient bien attentifs et consciencieux dans la réalisation de la tâche.

## 5 Conclusion

L'ergonomie des tâches est une problématique au coeur des plateformes de crowdsourcing. De même, le recueil de données fiables et pertinentes est aussi recherché avec un plus grand pouvoir d'expression. Nous travaillons sur la définition d'une nouvelle interface ayant pour objectif de solutionner ces problèmes. Dans cette étude la première phase de notre protocole expérimental est abordée. Ainsi, nous proposons une interface permettant au contributeur de renseigner sa certitude en sa réponse.

Les tests de cette interface sont effectués sur des données réelles provenant du laboratoire pour certaines et d'une campagne de crowdsourcing pour d'autres. Une analyse détaillée des résultats montre que plus la tâche est simple plus le taux de bonnes réponses des contributeurs est important. De même, la certitude du contributeur augmente lorsque la tâche devient plus simple. A l'inverse, le temps de réponse des contributeurs décroît avec la difficulté de la tâche. De plus, d'après une analyse en composante principale, le taux de bonnes réponses, la difficulté de la tâche et la certitude renseignée par le contributeur sont fortement positivement corrélées,

$\delta$	0	300	600	1200	1800	2400
MV	5	15	15	15	15	15
Probabilité	4	15	15	15	15	15
Fonctions de croyances	5	15	15	15	15	15

TAB. 3 – Nombre de bonnes réponses (/15) pour  $\Delta_1$ 

et négativement corrélées au temps de réponse du contributeur. Toujours d’après cette analyse, la perception de la difficulté de la tâche par le contributeur et sa capacité à renseigner sa certitude sont très fortement et positivement corrélés, et totalement indépendantes des autres variables énoncées précédemment. Il est également constaté d’après les résultats que le taux de bonnes réponses des contributeurs varie faiblement au début de l’expérience, puis se stabilise.

Nous avons modélisé les réponses des contributeurs de trois manières différentes : par vote majoritaire, par une approche probabiliste et par la théorie des fonctions de croyances. Il était supposé que la théorie des fonctions de croyance était l’approche la plus pertinente pour la modélisation des réponses mais ce n’est pas le cas ici puisque les trois méthodes comparées offrent les mêmes résultats. Ceci signifie que la majorité des contributeurs a raison, ceux-ci ont donc été attentifs et consciencieux dans la réalisation de la campagne de crowdsourcing.

Nos prochains travaux porteront sur la réalisation des deuxième et troisième phases expérimentales pour la définition de notre interface, portant respectivement sur l’imprécision, et l’incertitude combinée à l’imprécision.

## Remerciements

Ce travail est en partie financé par le conseil départemental des Côtes d’Armor et a été réalisé dans le cadre d’un Projet Exploratoire Premier Soutien (PEPS).

## Références

- Abassi, L. et I. Boukhris (2018). A worker clustering-based approach of label aggregation under the belief function theory. *Applied Intelligence*, 1–10.
- Ben Rjab, A., M. Kharoune, Z. Miklos, et M. Arnaud (2016). Characterization of experts in crowdsourcing platforms. *Belief Functions : Theory and Applications*. 9861.
- Dempster, A. P. (1967). Upper and lower probabilities induced by a multivalued mapping. *The Annals of Mathematical Statistics* 38, 325–339.
- Koulougli, D., A. Hadjali, et I. Rassoul (2016). Handling query answering in crowdsourcing systems : A belief function-based approach. In *Fuzzy Information Processing Society (NAFIPS), 2016 Annual Conference of the North American*, pp. 1–6. IEEE.
- Ouni, H., A. Martin, L. Gros, M. Kharoune, et Z. Miklos (2017). Une mesure d’expertise pour le crowdsourcing. *Extraction et Gestion des connaissances (EGC)*.

## Summary

Crowdsourcing is defined as the outsourcing of tasks to a crowd of contributors. The crowd is very diverse on these platforms and includes contributors with different skills and achieving the tasks more or less conscientiously. This profile diversity leads to uneven quality in the data. In order to overcome this problem we study a user interface offering the contributors a greater capacity of expression in their response. It allows the contributors to give imprecise answer in case of doubt and notified their confidence in their contributions. This interface was deployed asking users to answer at questions of variable complexity, for which ones these answers are known. Two range of complexity were tested and our results suggest that the user certainty depending on the absolute complexity of the question.